Biclustering Based on FCA and Partition Pattern Structures for Recommendation Systems

Nyoman Juniarta, Victor Codocedo, Miguel Couceiro, Amedeo Napoli

Biclustering Based on FCA and Partition Pattern Structures for Recommendation Systems

Nyoman Juniarta*, Victor Codocedo**
Miguel Couceiro*, Amedeo Napoli*

*Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
{nyoman.juniarta, miguel.couceiro, amedeo.napoli}@loria.fr
**Universidad Técnica Federico Santa María, Santiago, Chile
victor.codocedo@inria.cl

Abstract. This paper focuses on item recommendation for visitors in a museum within the framework of European Project CrossCult about cultural heritage. We present a theoretical research work about recommendation using biclustering. Our approach is based on biclustering using FCA and partition pattern structures. We investigate the possibility of incorporating the order information using this approach. Then, given the dataset of visitor trajectories, the result of our biclustering can be used to build a collaborative recommendation system.

1 Introduction

CrossCult (`http://www.crosscult.eu`) is a European project whose idea is to support the emergence of a European cultural heritage by allowing visitors in different cultural sites (e.g. museum, historic city, archaeological site) to improve the quality of their visit by using adapted computer-based devices. Such improvement can be accomplished by studying, among others, the possibility to build a dynamic recommendation system.

Here, our objective is to study a dynamic recommendation system for visitors in a museum (Hecht Museum for the case study). Given a new visitor V_n, the task is to suggest a museum item that may be interesting for him/her. Based on how a suggestion is made for a V_n, a recommendation system can be classified into one of the three categories Adomavicius and Tuzhilin (2005):

— *Content-based recommendations*: The system makes a suggestion based on the previous visited items of V_n.

— *Collaborative recommendations*: The system looks for previous users who have similar interest with V_n, and makes a suggestion based on their visited items.

— *Hybrid approaches*: The combination of content-based and collaborative approaches.

Our method belongs to the second category (collaborative recommendation). First we group all previous users based on their visit trajectories using biclustering. When V_n arrives, we try to find a G_s, i.e. a group who shares a similar interest to V_n. Then, based on the behavior of the visitors in G_s, we can suggest one item that may be interesting for V_n.

An approach using partition pattern structures was proposed to obtain biclusters with constant (or similar) values on the columns Codocedo and Napoli (2014). Our main contribution

1	1	4	3	5
1	1	2	5	1
3	3	4	2	1

TAB. 1 – *A bicluster (shaded) with constant value.*

(a)

4	2	5	3
4	2	5	3
4	2	5	3
4	2	5	3

(b)

1	1	1	1
2	2	2	2
4	4	4	4
3	3	3	3

(c)

4	2	5	3
3	1	4	2
5	3	6	4
7	5	8	6

(d)

1	2	4	3
3	5	7	6
2	3	8	4
4	5	9	8

(e)

1	2	4	3
0	1	1	2
5	4	6	4
6	5	7	5

TAB. 2 – *Examples of some bicluster types. (a) Constant columns, (b) constant rows, (c) additive coherent values, (d) coherent evolution on the columns, and (e) coherent evolution on the rows.*

here is to extend this approach to mine another type of biclusters: those with coherent evolution on the columns (CEC biclusters). We then use them to build a recommendation system. We focus on this bicluster type since we are dealing with the dataset of trajectories where each trajectory corresponds to an ordered list of items. Hence we are interested to take into account this ordering information, and CEC biclustering is one appropriate way to do that.

2 Biclustering

We consider a dataset composed of a set of objects, each of which has values over a set of attributes. This dataset can be represented as a numerical matrix C, where each entry c_{ij} indicates the value of object i w.r.t attribute j. One may be interested in finding which subset of objects possesses the same values w.r.t. a subset of attributes. Regarding the matrix representation, this is equivalent to the problem of finding a submatrix c that has a constant value over all of its elements (example in Table 1). This task is called biclustering with constant values, which is a simultaneous clustering of the rows and columns of a matrix.

Other than constant values, the bicluster approach also focused on finding other types of submatrices, as shown in Table 2. A bicluster with constant columns (rows) is a submatrix where each column (row) has the same value, as illustrated in Table 2a (Table 2b, resp.). In a bicluster with additive coherent values, the value of each cell c_{ij} follows the equation $\gamma + \alpha_i + \beta_j$, where γ is a constant, α_i is a constant value for row *i*, and β_j is a constant value for column *j*. For example, if $\gamma = 1$, $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (3, 2, 4, 6)$, and $(\beta_1, \beta_2, \beta_3, \beta_4) = (0, -2, 1, -1)$, then we can obtain the bicluster in Table 2c. Another interesting type is the CEC bicluster, also known as order-preserving submatrix Ben-Dor et al. (2003). In this type of bicluster, each row induces the same linear order across all columns. For example, in the bicluster in Table 2d, each row follows *column1* $<$ *column2* $<$ *column4* $<$ *column3*. Moreover, a bicluster with coherent evolution on the rows can be defined similarly, as shown in Table 2e.

Those different types of biclusters are useful when we are interested to identify a group of people who behave similarly according to a set of attributes. This group identification is necessary in the task of collaborative recommendation, because in the process of making a suggestion to a person, we first identify the people who are similar to him/her.

	i_1	i_2	i_3	i_4	i_5	i_6	i_7
v_1	1	2	3	4	5	6	7
v_2	2	4	5	3	7	1	6
v_3	4	2	1	5	6	3	7
v_4	7	3	1	4	2	6	5
v_a	?	?	1	2	?	?	?

TAB. 3 – *Order of visit of 7 items, observed from certain visitors.*

#	Visitors	Items (in order)
B1	v_1	$i_1, i_2, i_3, i_4, i_5, i_6, i_7$
B2	v_2	$i_6, i_1, i_4, i_2, i_3, i_7, i_5$
B3	v_1, v_3	i_3, i_4, i_5, i_7
B4	v_1, v_3	i_1, i_4, i_5, i_7
B5	v_1, v_4	i_3, i_5, i_6
B6	v_2, v_3	i_6, i_1, i_4, i_5
B7	v_1, v_2, v_3, v_4	i_2, i_7

TAB. 4 – *Some CEC biclusters in Table 3.*

3 Recommendation

In the context of CrossCult, we are working on a visitor dataset that comprises several trajectories in a museum. Within this project, our main objective is to build a dynamic recommendation system for new visitors. This system should be able to suggest a museum item to them based on their trajectory, by looking at the dataset of previous visitors. Also, it should be able to update the suggestion as they move inside the museum. Since we take into account the visit ordering of items when studying the trajectories, CEC biclustering seems to be a suitable approach. Here we will define some strategies for recommendation systems based on this type of biclustering.

Consider the dataset about 4 visitors in a museum with 7 items, given by Table 3. The numbers in row *x* indicate the path of visitor v_x. For example, the path of v_2 is $i_6 \rightarrow i_1 \rightarrow i_4 \rightarrow i_2 \rightarrow i_3 \rightarrow i_7 \rightarrow i_5$.

Now we have a new visitor, v_a, who recently arrives in the museum. She visits i_3, followed by i_4. Our task is to recommend a new item to her, by studying the CEC biclusters over the first four visitors. Some of those biclusters are listed in Table 4. From B7, we can see that all four visitors visit i_7 after (not before) i_2. Meanwhile in B3, regarding the items $\{i_3, i_4, i_5, i_7\}$, v_1 and v_3 follow the same order: $i_3 \rightarrow i_4 \rightarrow i_5 \rightarrow i_7$. These two visitors also agree in the order of items $\{i_1, i_4, i_5, i_7\}$, as seen in B4.

Those CEC biclusters can be studied to give a recommendation for v_a, by focusing on those who are similar to him/her. In the following subsections, we will present some possible strategies for performing this recommendation. They start from the same step: finding a CEC bicluster with a similar path to v_a.

This strategy focus on following sequential patterns in the dataset. The idea is if many people have the path $i_a \rightarrow i_b \rightarrow i_c$, then we should recommend the item i_c to a visitor who has done $i_a \rightarrow i_b$.

Since v_a has path $i_3 \rightarrow i_4$, we focus on the CEC biclusters that have those two items, i.e. B1, B2, and B3. One of those biclusters (B2) has a different ordering (i_3 after i_4), and thus we filter it out. Then, in B3 for example, the path is $i_3 \rightarrow i_4 \rightarrow i_5 \rightarrow i_7$. Therefore, we can recommend i_5 to v_a.

In this strategy, we assume that the order of visit reflects the order of interest. In this case, for $\mathtt{v}_2$, $\mathtt{i}_6$ is the most interesting (since (s)he visited it first) while $\mathtt{i}_5$ is the least interesting (since (s)he visited it last). Based on this assumption, if many people have the path $\mathtt{i}_a \rightarrow \mathtt{i}_b \rightarrow \mathtt{i}_c$ that reflects their interest, then we should recommend the item $\mathtt{i}_a$ to a visitor who has done $\mathtt{i}_b \rightarrow \mathtt{i}_c$, assuming that (s)he missed this item earlier in his/her visit.

Similar to the previous strategy, here we first identify CEC biclusters with similar path to $\mathtt{v}_a$. According to the visitors in B3, $\mathtt{i}_1$ and $\mathtt{i}_3$ are the most interesting items for $\mathtt{v}_1$ and $\mathtt{v}_3$ respectively. Therefore, we can suggest those two items to $\mathtt{v}_a$. Furthermore, we know that $\mathtt{v}_a$ has already visited $\mathtt{i}_3$, so we recommend her to visit $\mathtt{i}_1$.

In the framework of the CrossCult project, we are working on a specific dataset about the trajectories of 254 visitors in Hecht Museum in Haifa, Israel Lanir et al. (2013). In this dataset, there are 52 items over 8 rooms (A–G). Therefore, the visitor–item matrix for biclustering will have 254 rows and 52 columns.

4 Conclusion

In this work, we explore an approach to build collaborative recommendation strategy for visitors in a museum. This strategy takes into account the order of interest or the order of visit for each visitor, and we showed how to use CEC biclustering to obtain a set of similar visitors. Further comparisons of CEC biclustering and sequential pattern mining Yan et al. (2003); Yu et al. (2016) should be investigated, in particular, regarding their complexities and their results.

References

Adomavicius, G. and A. Tuzhilin (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering 17*(6), 734–749.

Alqadah, F., C. K. Reddy, J. Hu, and H. F. Alqadah (2015). Biclustering neighborhood-based collaborative filtering method for top-n recommender systems. *Knowledge and Information Systems 44*(2), 475–491.

Ben-Dor, A., B. Chor, R. Karp, and Z. Yakhini (2003). Discovering local structure in gene expression data: the order-preserving submatrix problem. *Journal of computational biology 10*(3-4), 373–384.

Codocedo, V. and A. Napoli (2014). Lattice-based biclustering using partition pattern structures. In *Proceedings of the Twenty-first European Conference on Artificial Intelligence*, pp. 213–218. IOS Press.

Lanir, J., T. Kuflik, E. Dim, A. J. Wecker, and O. Stock (2013). The influence of a location-aware mobile guide on museum visitors' behavior. *Interacting with Computers 25*(6), 443–460.

Yan, X., J. Han, and R. Afshar (2003). CloSpan: Mining: Closed sequential patterns in large datasets. In *Proceedings of the 2003 SIAM international conference on data mining*, pp. 166–177. SIAM.

Yu, F., Q. Liu, S. Wu, L. Wang, and T. Tan (2016). A dynamic recurrent model for next basket recommendation. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 729–732. ACM.